

AD-A190 323

## REPORT DOCUMENTATION PAGE

DTIC FILE COPY (2)

1a. SECURITY CLASSIFICATION AUTHORITY:		1b. RESTRICTIVE MARKINGS	
2a. DECLASSIFICATION/DOWNGRADING SCHEDULE		3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) Technical Report No. 206		5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR-87-1851	
6a. NAME OF PERFORMING ORGANIZATION University of North Carolina	6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION AFOSR/RM	
6c. ADDRESS (City, State, and ZIP Code) Statistics Dept. 321-A Phillips Hall X000CA CB 3260 Chapel Hill, NC 20508 27599-3260		7b. ADDRESS (City, State, and ZIP Code) AFOSR/RM Bldg 410 Bolling AFB DC 20332-8448	
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR	8b. OFFICE SYMBOL (If applicable) RM	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFOSR-No. F49620 85C 0144.	
8c. ADDRESS (City, State, and ZIP Code) AFOSR/RM Bldg 410 Bolling AFB DC 20332-8448		10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 61102F	PROJECT NO. 2304
		TASK NO. A-5	WORK UNIT ACCESSION NO.
11. Title (Include Security Classification) Estimation of convolution tails			
12. PERSONAL AUTHOR(S) Willekens, E.			
13a. TYPE OF REPORT Preprint	13b. TIME COVERED FROM 10/1/86 TO 9/30/87	14. DATE OF REPORT (Year, Month, Day) September 1987	15. PAGE COUNT 18
16. SUPPLEMENTARY NOTATION N/A			
17. COSATI CODES FIELD GROUP SUB-GROUP		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number) Key Words & Phrases: subexponential distributions, U-statistics strong consistency, asymptotic normality.	
19. ABSTRACT (Continue on reverse if necessary and identify by block number) Several classes of distribution functions (d.f.) are originated by considering distributions whose tailfunctions satisfy special asymptotic relations. A large class sharing this property is provided by the subexponential class, in which case the asymptotic relation involves tails of convolution powers. In this paper we introduce a statistic which estimates the asymptotic behaviour of convolution tails of a given d.f. and we show that this statistic is strongly consistent and asymptotically normal under appropriate conditions. Furthermore, the statistic can be used to test the hypothesis that a d.f. is in $\mathcal{V}$ .  it is shown the exponential class being			
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT <input checked="" type="checkbox"/> UNCLASSIFIED/UNLIMITED <input type="checkbox"/> SAME AS RPT. <input type="checkbox"/> DTIC USERS		21. ABSTRACT SECURITY CLASSIFICATION Unclassified/unlimited	
22a. NAME OF RESPONSIBLE INDIVIDUAL Maj. Brian Woodruff		22b. TELEPHONE (Include Area Code) (202) 767-5026/7	22c. OFFICE SYMBOL RM described.

DD FORM 1473, 84 MAR

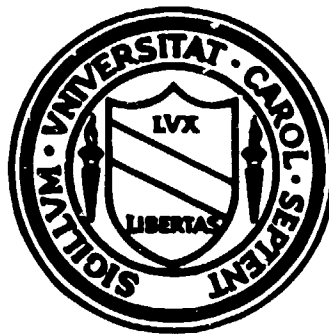
83 APR edition may be used until exhausted.  
All other editions are obsolete.SECURITY CLASSIFICATION OF THIS PAGE  
Unclassified/Unlimited

87 12 29 201

# **CENTER FOR STOCHASTIC PROCESSES**

**AOSR-TR- 87 - 1851**

**Department of Statistics  
University of North Carolina  
Chapel Hill, North Carolina**



**Estimation of convolution tails**

**by**

**Eric Willekens**

**Technical Report No. 206**

**September 1987**

**87 12 29 261**

# Estimation of convolution tails

Eric Willekens\*

Catholic University Leuven  
and

University of North Carolina at Chapel Hill

## Abstract:

Several classes of distribution functions (d.f.) are originated by considering distributions whose tailfunctions satisfy special asymptotic relations. A large class sharing this property is provided by the subexponential class  $\mathcal{S}$ , in which case the asymptotic relation involves tails of convolution powers. In this paper we introduce a statistic which estimates the asymptotic behaviour of convolution tails of a given d.f. and we show that this statistic is strongly consistent and asymptotically normal under appropriate conditions. Furthermore, the statistic can be used to test the hypothesis that a d.f. is in  $\mathcal{S}$ .

**Keywords and Phrases:** subexponential distributions, U-statistics, strong consistency, asymptotic normality.

\*Research Assistant of the Belgian National Fund for Scientific Research.

Research partially supported by Air Force Office of Scientific Research Grant No. F49620 85C 0144.



Accession For	
NTIS GRA&I	<input checked="checked" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

## 1. Introduction

We work with distribution functions (d.f.)  $F$  which satisfy  $F(0-)=0$  and  $F(x) < 1$  for every  $x \in \mathbb{R}$ . We use  $*$  to denote the usual convolution product, i.e. the convolution powers of  $F$  are given as  $F^{*2}(x) = \int_0^x F(x-y)dF(y)$ ,  $F^{*n} = F^{*(n-1)} * F$ ,  $n=1,2,\dots$  and  $F^{*0} = \delta_0$ , the d.f. of the unit mass at zero. Throughout the paper, the tailfunction of a given d.f.  $G$  will be denoted as  $\bar{G}$ , i.e.  $\bar{F} = 1-F$ ,  $\bar{F}^{*n} = 1-F^{*n}$  etc....

We now define the class  $\mathcal{V}$  of subexponential distributions.

### Definition 1.1

A d.f.  $F$  is called subexponential ( $F \in \mathcal{V}$ ) iff

$$(1.1) \quad \lim_{x \rightarrow \infty} \frac{\bar{F}^{*m}(x)}{\bar{F}(x)} = m \quad \text{for some integer } m \geq 2.$$

It is well known that if (1.1) holds, it holds for all integers  $m \geq 2$  (see [2]).

The class  $\mathcal{V}$  was introduced independently by Chiskyakov [2] and Chover et al. [3]. Both authors used this type of d.f. to model the lifetime distribution in a subcritical branching process in order to determine the asymptotic behaviour of the mean population size [2] [4].

Chover et al. [3] also introduced the class  $\mathcal{SD}$  of densities corresponding to the d.f. in definition 1.1

### Definition 1.2.

A probability density  $f > 0$  is subexponential ( $f \in \mathcal{SD}$ )

$$(1.2) \quad (i) \quad \lim_{x \rightarrow \infty} \frac{f(x-y)}{f(x)} = 1 \quad \text{for every } y \in \mathbb{R}$$

$$(1.3) \quad (ii) \quad \lim_{x \rightarrow \infty} \frac{f^{*2}(x)}{f(x)} = 2$$

Here  $\times$  denotes density convolution, i.e.  $f \times g = \int_0^x f(x-y)g(y)dy$  and should not be confused with  $*$ .

It is clear by de l'Hopitals' theorem that if a d.f.  $F$  has a density  $f \in \mathcal{SD}$ ,

also  $F \in \mathcal{V}$ .

Ever since they were originated, the classes  $\mathcal{V}$  and SD have been studied extensively by a number of authors, such as Teugels [2], Pitman [17], Embrechts and Goldie [8], [9], Cline [6], Omey and Willekens [14] [15]. Since subexponential d.f. characterize a certain tail behaviour of compound distributions, applications of the previously defined classes may be found in various domains of stochastic processes, see e.g. [10] [12], [24] and references therein.

Because of the convolution power in definition 1.1 it is often very hard to check if a given d.f. satisfies (1.1). Sufficient conditions for membership of  $\mathcal{V}$  only in terms of the tail of  $F$  are known [6] [11], but require an analytical expression for  $\bar{F}$ . If a d.f. is only known through a finite number of observations, it is impossible with the present theory to decide whether this d.f. is subexponential or not. Such situations frequently arise in some applied stochastic models such as queuing and risk theory, see [13].

In this paper, we are concerned with developing a statistical approach to subexponentiality, in the sense that we want to define a statistic, based on a given sample  $X_1, X_2, \dots, X_n$ , which gives us valuable information to decide whether the underlying d.f. is subexponential or not.

In the next section, we introduce a statistic which reflects the subexponential property of its corresponding distribution. In section 3 we prove that the statistic is strongly consistent while section 4 is devoted to the asymptotic normality. Finally section 5 contains some comments and concluding remarks.

## 2. Definition of the statistic

Let  $X_1, X_2, \dots, X_n$  be a sequence of independent identically distributed (i.i.d.) random variables with d.f.  $F$ , and denote by  $X_{1:n} \leq \dots \leq X_{n:n}$  the order statistics of the sample. The following statistic is the sample version of (1.1) and seems a plausible choice to describe subexponentiality:

$$H_n(x) := \frac{\overline{F}_n^{*m}(x)}{\overline{F}_n(x)} \cdot I_{\Lambda_n}(x).$$

Here  $\overline{F}_n$  denotes the empirical d.f. based on  $X_1, \dots, X_n$  and  $\Lambda_n(x) = \{\omega: \exists i: 1 \leq i \leq n: X_i(\omega) > x\}$ .

Notice that the indicator function  $I_{\Lambda_n}$  is necessary to make sure that  $H_n(x)$  is well defined.

Since  $H_n(x)$  is a ratio of V-statistics (see [20]), and  $I_{\Lambda_n}(x) \rightarrow 1$  almost surely (a.s.) for fixed  $x$ , we get that

$$\lim_{n \rightarrow \infty} H_n(x) = \frac{\overline{F}^{*m}(x)}{\overline{F}(x)} \quad \text{a.s.} \quad \text{for fixed } x.$$

In order that  $H_n(x)$  can give a meaningful description of the subexponential behaviour of  $F$ , we have to let  $x \rightarrow \infty$ , which gives

$$\lim_{x \rightarrow \infty} \lim_{n \rightarrow \infty} H_n(x) = 2 \quad \text{iff} \quad F \in \mathcal{G}.$$

This relation shows that  $H_n$  contains information about  $\mathcal{G}$ , but, is in a sense useless because of the two limits. One way to solve this problem is to substitute for  $x$  a deterministic sequence  $(x_n)_n$  and letting  $n \rightarrow \infty$ . In this case however, the remaining statistic depends on a parameter which has to be chosen artificially and it turns out that this parameter rather heavily depends on the d.f.  $F$ , which is generally unknown (see [24]).

We therefore propose to replace the deterministic sequence by a random sequence by taking for each  $n$  one of the observations, which almost surely tends to infinity as  $n \rightarrow \infty$ .

If we choose for each  $n$  the intermediate order statistic  $X_{n-k_n:n}$  with  $k_n \rightarrow \infty$  but  $\frac{k_n}{n} \rightarrow 0$ , then  $H_n(X_{n-k_n:n})$  reduces to

$$H_n(X_{n-k_n:n}) = \frac{1}{k_n \cdot n^{m-1}} \sum_{c'} I(X_{i_1} + X_{i_2} + \dots + X_{i_m}, X_{n-k_n:n})$$

where  $\sum_{c'}$  denotes the summation over all  $m$ -tuples  $(i_1, \dots, i_m)$  consisting of  $m$  elements of  $\{1, \dots, n\}$ .

We now slightly modify  $H_n(X_{n-k_n:n})$  by removing the sum over all  $m$ -tuples which at least contain two equal integers. This will not affect the asymptotic behaviour because their contribution to the whole sum is of a smaller magnitude than the sum of the remaining terms. Finally changing the normalizing constant a little, we end up with the statistic we will discuss in the next sections:

$$(2.1) \quad U_n(X_{n-k_n:n}) := \frac{n}{k_n} \frac{1}{\binom{n}{m}} \sum_c I(X_{i_1} + X_{i_2} + \dots + X_{i_m} > X_{n-k_n:n}).$$

Here  $\sum_c$  stands for the sum over all combinations of  $m$  distinct elements out of  $\{1, \dots, n\}$ .

Clearly for each fixed  $x$ ,  $\frac{k_n}{n} U_n(x)$  is a  $U$ -statistic and it seems irresistible to use the well known asymptotic theory for  $U$ -statistics [20] in order to determine the behaviour of  $U_n(X_{n-k_n:n})$ .

However the presence of  $(X_{n-k_n:n})$  makes the kernel stochastic and  $n$ -dependent,

so that  $U_n(X_{n-k_n:n})$  is in fact a  $U$ -statistic with an estimated parameter, see

[16], [18]. Asymptotic normality of such statistics was studied among others by

Randles [18]. His method however is only valid when the estimated parameter is

constant, while in our case  $X_{n-k_n:n}$  is (under the appropriate conditions) a

consistent estimator for  $x_n$ , where  $x_n$  is the intermediate population quantile,

given by the equation  $\bar{F}(x_n) = \frac{k_n}{n}$ . And since  $\frac{k_n}{n} \rightarrow 0$ ,  $x_n \rightarrow \infty$ .

In the next sections we adapt and modify Randles method to make it work in our

case. The basic tool in establishing this is an a.s. Bahadur representation for

$X_{n-k_n:n}$ . This is provided by Watts [23]:

let  $F$  be a d.f. such that on an interval  $(c, \infty)$ ,  $c > 0$ ,  $F'(x) = f(x)$  and  $f'(x)$  exist with  $f(x) > 0$ , and suppose that there is a constant  $M$  and a function  $V$  with  $V(x) = O(x)$  ( $x \rightarrow \infty$ ) such that

$$(2.2) \quad \frac{\bar{F}(x)}{V(x)f(x)} < M \quad \text{and} \quad \frac{\bar{F}(x)|f'(x+y)|}{f^2(x)} < M$$

for large  $x$ , where  $y = o(V(x))$  ( $x \rightarrow \infty$ ). Define

$$L(x, y) := \sup\{f(z) \mid |z-x| \leq y\} \quad \text{and} \quad M(x, y) := \inf\{f(z) \mid |z-x| \leq y\}$$

and assume whenever  $y = o(V(x))$ ,

$$(2.3) \quad \frac{L(x, y)}{f(x)} \quad \text{and} \quad \frac{f(x)}{M(x, y)} \quad \text{remain bounded as } x \rightarrow \infty.$$

Then, under (2.2) and (2.3), for any sequence  $k_n$  with  $\frac{k_n}{n} \rightarrow 0$  and  $\frac{k_n}{\log^3 n} \rightarrow \infty$ ,

$$(2.4) \quad X_{n-k_n:n} = x_n + \frac{n \bar{F}_n(x_n) - k_n}{nf(x_n)} + R_n$$

where  $\bar{F}_n(x_n) = \frac{k_n}{n}$  and

$$(2.5) \quad R_n = O((k_n \log^3 n)^{1/4} / nf(x_n)) \quad \text{as } n \rightarrow \infty \quad \text{with probability one.}$$

Before we proceed investigating the properties of  $U_n(X_{n-k_n:n})$ , we first show

that the conditions (2.2) and (2.3) are satisfied for a large subclass of subexponential distributions. Also notice that every d.f.  $F$  in  $\mathcal{V}$  is close to a subexponential d.f.  $G$  (in the sense that  $\bar{F}(x) \sim \bar{G}(x)$  ( $x \rightarrow \infty$ )) with  $G$  infinitely many times differentiable [19], so that it is no loss of generality to assume that the derivatives of  $F$  exist.

We need the following class of functions which generalizes properly the class of slowly varying functions (see [1]):  $L \in \text{OSV}$  iff there exist absolute constants  $0 < c \leq C < \infty$  such that for every  $t \geq 1$ ,

$$0 < c \leq \liminf_{x \rightarrow \infty} \frac{L(xt)}{L(x)} \leq \limsup_{x \rightarrow \infty} \frac{L(xt)}{L(x)} \leq C < \infty.$$



Theorem 2.1

Let  $F \in \mathcal{V}$  and  $L \in \text{OSV}$  such that

$$(2.6) \quad \bar{F}(x) = \exp - x^\alpha \int_1^x \frac{L(y)}{y} dy \quad (0 \leq \alpha \leq 1).$$

Assume that

$$(i) \quad \limsup_{x \rightarrow \infty} \left| \frac{L'(x) \cdot x^{1-\alpha}}{\left(\int_1^x \frac{L(y)}{y} dy\right)^2} \right| < \infty \quad \text{if } \alpha > 0$$

$$(ii) \quad L \text{ is bounded away from 0 and } \limsup_{x \rightarrow \infty} |xL'(x)| < \infty \quad \text{if } \alpha = 0.$$

Then (2.2) and (2.3) are satisfied.

Proof.

$$(2.7) \quad \begin{aligned} \text{Since } V_\alpha^{-1} &:= \frac{f(x)}{\bar{F}(x)} = x^{\alpha-1} \left( \alpha \int_1^x \frac{L(y)}{y} dy + L(x) \right) \sim \alpha x^{\alpha-1} \int_1^x \frac{L(y)}{y} dy \quad \text{if } \alpha > 0 \\ &= x^{-1} L(x) \quad \text{if } \alpha = 0 \end{aligned}$$

it is clear that the first property in (2.2) is satisfied.

First consider  $\alpha > 0$ : with the expression for  $V_\alpha$  in (2.7), it is easy to see that  $\bar{F}(x + o(V_\alpha(x))) \sim \bar{F}(x)$  ( $x \rightarrow \infty$ ) and then (2.3) follows immediately from the regular variation of  $V_\alpha$  and (2.7).

To show the second part of (2.2), it is sufficient to prove that  $\left| \left( \frac{\bar{F}}{F} \right)' \right|$  is bounded, which is satisfied by (i).

If  $\alpha = 0$ ,  $V_0(x) = \frac{x}{L(x)} = O(x)$  since  $L$  is bounded away from zero.

The same reasoning as for the case  $\alpha > 0$  completes the proof.  $\square$

Remarks

1. Clearly the representation in (2.6) contains all important archetypes of subexponential d.f. However, since  $L$  is bounded from below if  $\alpha = 0$ , it always implies that

$$(2.8) \quad \bar{F}(x) = o(x^{-\epsilon}) \quad (x \rightarrow \infty)$$

for some  $\epsilon > 0$ , such that (2.6) does not cover d.f. with slowly varying tails.

This is not surprising as such distributions generally violate the first

condition in (2.2), see e.g. [7].

2. Clearly (2.3) implies that  $\bar{F}(x + o(V(x))) = O(\bar{F}(x))$  ( $x \rightarrow \infty$ ), whereas the representation in (2.6) gives that

$$(2.9) \quad \bar{F}(x + o(V(x))) \sim \bar{F}(x) \quad (x \rightarrow \infty).$$

This somewhat stronger condition will be used in section 3.

### 3. Strong consistency of $U_n(X_{n-k_n:n})$

Theorem 3.1. Let  $X_1, X_2, \dots, X_n$  be a sequence of i.i.d. random variables with d.f.  $F$  and let  $F' = f$  a SD. Suppose that (2.2), (2.3), (2.8) and (2.9) are satisfied.

If  $(k_n)_n$  is a regularly varying sequence such that  $\frac{k_n}{n} \rightarrow 0$  and  $k_n / \log^3 n \rightarrow \infty$ , then

$$U_n(X_{n-k_n:n}) \rightarrow m \quad \text{a.s.}$$

Proof. From the a.s. Bahadur representation in (2.4) and the law of iterated logarithm for triangular arrays [5], we easily obtain that

$$X_{n-k_n:n} \stackrel{\text{a.s.}}{=} x_n + T_n + R_n := h(n)$$

$$\text{where } T_n \stackrel{\text{a.s.}}{=} O\left[\frac{\sqrt{k_n \log \log k_n / n f(x_n)}}{n f(x_n)}\right] \quad (n \rightarrow \infty).$$

and  $R_n$  is as in (2.5).

It is therefore sufficient to show that

$$U_n(h(n)) \rightarrow m \quad \text{a.s.} \quad \text{as } n \rightarrow \infty.$$

Let  $\frac{k_n}{n} \tilde{U}_n(x)$  denote the projection of the U-statistic  $\frac{k_n}{n} U_n(x)$  on the basic observations ([20, p. 187]), then one easily calculates with the method of [20, p. 182] that for every  $n \geq m$ ,

$$(3.1) \quad \left(\frac{k_n}{n}\right)^2 \binom{n}{2} E\{\tilde{U}_n(x) - U_n(x)\}^2 = (\bar{F}^{(n)}(x))^2 + \{\bar{F}^{(n)}(x) - n\bar{F}(x)\} \\ - n \int_0^x (\bar{F}^{(n-1)}(x-y))^2 dF(y) = \\ o(\bar{F}(x)) \quad (x \rightarrow \infty).$$

The last step follows from the fact that  $f$  is SD and [14, lemma 3.1.1]. From the first condition in (2.2), we have for  $n$  sufficiently large that

$$\frac{\sqrt{k_n}}{nf(x_n)} \leq \frac{M \cdot V(x_n)}{\sqrt{k_n}}.$$

Furthermore, (2.8) implies that  $x_n = o\left[\left(\frac{n}{k_n}\right)^{1/4}\right]$  ( $n \rightarrow \infty$ ), such that

$$\frac{T_n}{V(x_n)} = o\left[\left[\frac{\log(1/e \log n/k_n)}{k_n}\right]^{1/2}\right] \quad (n \rightarrow \infty).$$

Since  $k_n/\log^3 n \rightarrow \infty$ , we have that  $T_n = o(V(x_n))$  ( $n \rightarrow \infty$ ), and since  $R_n = o(T_n)$  ( $n \rightarrow \infty$ ), we obtain that

$$(3.2) \quad h(n) = x_n + o(V(x_n)) \quad (n \rightarrow \infty).$$

Together with (3.1) and (2.9), this implies that

$$E\{\tilde{U}_n(h(n)) - U_n(h(n))\}^2 = o\left[\frac{\bar{F}(h(n))}{k_n^2}\right] = o\left[\frac{1}{nk_n}\right] \quad (n \rightarrow \infty)$$

and since  $\log^3 n = o(k_n)$  ( $n \rightarrow \infty$ ), it follows from Chebychev's inequality and the Borel-Cantelli lemma that

$$(3.3) \quad \tilde{U}_n(h(n)) - U_n(h(n)) \rightarrow 0 \quad \text{a.s.} \quad (n \rightarrow \infty).$$

From (3.2) and (2.9) we have that  $\bar{F}^{(n)}(h(n)) \sim n\bar{F}(x_n)$  ( $n \rightarrow \infty$ ) such that with (3.3), the proof is finished if we can show that

$$(3.4) \quad \tilde{U}_n(h(n)) - \frac{n}{k_n} \bar{F}^{(n)}(h(n)) \rightarrow 0 \quad \text{a.s.} \quad (n \rightarrow \infty).$$

Now

$$(3.5) \quad \frac{k_n}{n} \tilde{U}_n(x) = \frac{n}{n} \sum_{i=1}^n \bar{F}^{(n-1)}(x - X_i) - (n-1) \bar{F}^{(n)}(x)$$

so that

$$\begin{aligned}
& E\{\tilde{U}_n(h(n)) - \frac{n}{k_n} \bar{F}^{*m}(h(n))\}^2 \\
&= \frac{nm^2}{k_n^2} \left\{ \int_0^\infty (\bar{F}^{*m-1}(h(n) - y))^2 dF(y) - (\bar{F}^{*m}(h(n)))^2 \right\} \\
(3.6) \quad & \sim \frac{n \cdot m^2}{k_n^2} \bar{F}(h(n)) \quad (n \rightarrow \infty) \\
& \sim \frac{m^2}{k_n} \quad (n \rightarrow \infty).
\end{aligned}$$

As in (3.1), (3.6) follows from the fact that  $f \in SD$  and [14, lemma 3.1.1].

Taking  $\sigma > 1$  arbitrary, and putting  $n_\ell = [\sigma^\ell]$ ,  $\ell=1,2,\dots$ , it follows from (3.6) and the Borel-Cantelli lemma that (3.4) holds if the limit is taken over the subsequence  $(n_\ell)_\ell$ . Our aim however is to let  $\sigma \rightarrow 1$  and to prove convergence over the whole sequence.

Take  $n > 0$  arbitrary and let  $\ell = \ell(\sigma, n)$  be such that  $[\sigma^\ell] \leq n < [\sigma^{\ell+1}]$ . Without loss of generality we may assume that the sequence  $(h(n))_n$  is monotone non decreasing, such that

$$(3.7) \quad I_{1,\sigma}(\ell) \leq \tilde{U}_n(h(n)) - \frac{n}{k_n} \bar{F}^{*m}(h(n)) \leq I_{2,\sigma}(\ell)$$

with

$$I_{2,\sigma}(\ell) = \frac{m}{k_n} \sum_{i=1}^n \{ \bar{F}^{*m-1}(h([\sigma^\ell]) - X_i) - \bar{F}^{*m}(h([\sigma^{\ell+1}])) \}$$

and

$$I_{1,\sigma}(\ell) = \frac{m}{k_n} \sum_{i=1}^n \{ \bar{F}^{*m-1}(h([\sigma^{\ell+1}]) - X_i) - \bar{F}^{*m}(h([\sigma^\ell])) \}.$$

Now

$$\begin{aligned}
I_{2,\sigma}(\ell) &= \frac{m}{k_n} \sum_{i=1}^n \{ \bar{F}^{*m-1}(h([\sigma^\ell]) - X_i) - \bar{F}^{*m}(h([\sigma^\ell])) \} \\
&\quad + \frac{m \cdot n}{k_n} \{ \bar{F}^{*m}(h([\sigma^\ell])) - \bar{F}^{*m}(h([\sigma^{\ell+1}])) \}.
\end{aligned}$$

Clearly the first term in the right hand side tends a.s. to zero as  $\ell \rightarrow \infty$ .

Since  $f \in SD$ , the second term is asymptotically equal to

$$\frac{m \cdot n}{k_n} \{ \bar{F}(h([\sigma^\ell])) - \bar{F}(h([\sigma^{\ell+1}])) \} \sim m^2 \{ 1 - k_{[\sigma^{\ell+1}]} \sigma^{-1} k_{[\sigma^\ell]}^{-1} \} \frac{n}{k_n} \cdot \frac{k_{[\sigma^\ell]}}{[\sigma^\ell]}.$$

Since  $(k_n)_n$  is a regularly varying sequence, we then have that

$$\lim_{\sigma \rightarrow 1} \lim_{\ell \rightarrow \infty} I_{2,\sigma}(\ell) = 0.$$

The same treatment for  $I_{1,\sigma}(\ell)$  and (3.7) imply (3.4). This completes the proof.  $\square$

#### 4. Asymptotic normality of $U_n(X_{n-k_n:n})$ .

For proving asymptotic normality, we use the following smoothness condition which is somewhat stronger than  $f \in \text{SD}$ :

$$(4.1) \quad \lim_{x \rightarrow \infty} \frac{(f^{x2}(x))'}{f'(x)} = 2.$$

Clearly from de l'Hôpital's theorem, (4.1) implies  $f \in \text{SD}$ .

#### Theorem 4.1.

Let  $X_1, X_2, \dots, X_n$  be a sequence of i.i.d. random variables with d.f.  $F$  and let  $F' = f$ . Suppose that (2.2), (2.3) and (4.1) are satisfied. If  $(k_n)_n$  is a sequence such that  $\frac{k_n}{n} \rightarrow 0$  and  $k_n / \log^3 n \rightarrow \infty$  and if  $\bar{F}(x_n) = k_n/n$ , then

$$\sqrt{k_n} \left[ U_n(X_{n-k_n:n}) - \frac{\bar{F}^{xm}(x_n)}{\bar{F}(x_n)} \right] \xrightarrow{d} N(0, 4m^2)$$

#### Proof

For the proof of the theorem we follow Randles [18] and we split up

$$\frac{k_n}{n} U_n(X_{n-k_n:n}) - \bar{F}^{xm}(x_n) = A_1(m, n) + A_2(m, n) \text{ where}$$

$$(4.2) \quad A_1(m, n) = \frac{k_n}{n} U_n(x_n) - \bar{F}^{xm}(x_n) + \bar{F}^{xm}(X_{n-k_n:n}) - \bar{F}^{xm}(x_n)$$

and

$$(4.3) \quad A_2(m, n) = \frac{k_n}{n} U_n(X_{n-k_n:n}) - \bar{F}^{xm}(X_{n-k_n:n}) - \frac{k_n}{n} U_n(x_n) + \bar{F}^{xm}(x_n).$$

Each of the terms  $A_1(m, n)$  and  $A_2(m, n)$  will be considered in a separate lemma, and the proof of theorem 4.1 follows immediately from a combination of both.

Lemma 4.2

Under the conditions of theorem 4.1,

$$\frac{n}{\sqrt{k_n}} A_1(m, n) \xrightarrow{d} N(0, 4m^2)$$

Proof

Using a two term Taylor expansion, we can write

$$(4.4) \quad \bar{F}^{xm}(X_{n-k_n:n}) - \bar{F}^{xm}(x_n) = (X_{n-k_n:n} - x_n) f^{xm}(x_n) + \frac{1}{2}(X_{n-k_n:n} - x_n)^2 (f^{xm})'(\theta_n)$$

where  $|\theta_n(\omega) - x_n| \leq |X_{n-k_n:n}(\omega) - x_n|$ . The Bahadur representation in (2.4)

implies that

$$(4.5) \quad \frac{nf(x_n)}{\sqrt{k_n}} (X_{n-k_n:n} - x_n) = O_p(1)$$

such that  $\theta_n = x_n + O_p\left(\frac{\sqrt{k_n}}{nf(x_n)}\right) \quad (n \rightarrow \infty)$ .

Clearly by (2.2),  $\frac{\sqrt{k_n}}{nf(x_n)} = o(V(x_n)) \quad (n \rightarrow \infty)$ .

Furthermore, from (4.1) and [14, lemma 3.1.3], we have that

$$\lim_{x \rightarrow \infty} \frac{(f^{xm})'(x)}{f'(x)} = m + 1.$$

Using this, we can write

$$(4.6) \quad \frac{1}{2}(X_{n-k_n:n} - x_n)^2 (f^{xm})'(\theta_n) = f'(x_n + O_p(\frac{\sqrt{k_n}}{nf(x_n)})) O_p(\frac{k_n}{n^2 f^2(x_n)}) \quad (n \rightarrow \infty).$$

With  $\tilde{U}_n$  defined as in (3.5), we have from (3.1) that

$$(4.7) \quad U_L(x_n) = U(x_n) + o_p\left(\frac{\bar{F}(x_n)^{1/2}}{k_n}\right) \quad (n \rightarrow \infty).$$

Combination of (4.4) - (4.7), (3.5) and (2.4) implies that

$$\begin{aligned}
(4.8) \quad \frac{n}{\sqrt{k_n}} A_1(m, n) &= \frac{m}{\sqrt{k_n}} \sum_{i=1}^n \{ \bar{F}^{xm-1}(x_n - X_i) - \bar{F}^{xm}(x_n) + I(X_i > x_n) - \bar{F}(x_n) \} \\
&\quad + \left( \frac{f^{xm}(x_n)}{f(x_n)} - m \right) \cdot o_p(1) + \frac{f^{xm}(x_n)}{f(x_n)} O\left(\left(\frac{\log^3 n}{k_n}\right)^{1/4}\right) \\
&\quad + \frac{f'(x_n + o_p(\frac{\sqrt{k_n}}{nf(x_n)})) \bar{F}(x_n)}{f^2(x_n)} \cdot o_p\left(\frac{1}{\sqrt{k_n}}\right) + o_p\left(\left(\frac{\bar{F}(x_n)}{k_n}\right)^{1/2}\right) \quad (n \rightarrow \infty).
\end{aligned}$$

Since

$$\begin{aligned}
\text{Var}\{\bar{F}^{xm-1}(x_n - X_i) + I(X_i > x_n)\} \\
&= \int_0^\infty (\bar{F}^{xm-1}(x_n - y))^2 dF(y) + 3\bar{F}(x_n) - (\bar{F}^{xm}(x_n) + \bar{F}(x_n))^2 \\
&\sim \bar{F}(x_n) \quad (1+3) \\
&= \frac{4 \cdot k_n}{n} \quad (n \rightarrow \infty).
\end{aligned}$$

and since all remainder terms in (4.8) tend to zero, it is easy to see by the central limit theorem for triangular arrays [5] that the desired limit law for  $A_1(m, n)$  holds. This completes the proof.  $\square$

#### Lemma 4.3.

Under the conditions of theorem 4.1,

$$(4.9) \quad \frac{n}{\sqrt{k_n}} A_2(m, n) \xrightarrow{p} 0.$$

#### Proof

Denote

$$h(x_{i_1}, x_{i_2}, \dots, x_{i_m}; q) := I(x_{i_1} + x_{i_2} + \dots + x_{i_m} > q) - \bar{F}^{xm}(q)$$

and put

$$Q_n(s) :=$$

$$\frac{n}{\sqrt{k_n}} \frac{1}{\binom{n}{m}} \sum_c \{h(x_{i_1}, x_{i_2}, \dots, x_{i_m}; x_n + \frac{\sqrt{k_n}}{nf(x_n)} s) - h(x_{i_1}, x_{i_2}, \dots, x_{i_m}; x_n)\}.$$

In this notation, (4.9) is equivalent to showing that

$$Q_n\left(\frac{nf(x_n)}{\sqrt{k_n}} (X_{n-k_n:n} - x_n)\right) \xrightarrow{p} 0.$$

By (4.5), it is therefore enough to prove that for some bounded interval C,

$$(4.10) \quad P(\sup_{s \in C} Q_n(s) > \epsilon) \rightarrow 0 \quad (n \rightarrow \infty),$$

for every  $\epsilon > 0$ .

The way to show (4.10) follows more or less the same lines as the proof of [21, Theorem 3.1].

We first investigate the differences of the kernel  $h$ : for  $0 \leq s < t$ ,

$$\begin{aligned} (4.11) \quad & E|h(X_{i_1}, X_{i_2}, \dots, X_{i_m}; x_n + \frac{\sqrt{k_n}}{nf(x_n)} \cdot t) - h(X_{i_1}, X_{i_2}, \dots, X_{i_m}; x_n + \frac{\sqrt{k_n}}{nf(x_n)} \cdot s)| \\ & \leq E\{I(x_n + \frac{\sqrt{k_n}}{nf(x_n)} s \leq X_{i_1} + \dots + X_{i_m} \leq x_n + \frac{\sqrt{k_n}}{nf(x_n)} t) + F^{*m}(x_n + \frac{\sqrt{k_n}}{nf(x_n)} \cdot t) \\ & \quad - F^{*m}(x_n + \frac{\sqrt{k_n}}{nf(x_n)} s)\} = 2 \cdot \frac{\sqrt{k_n}}{nf(x_n)} \int_s^t f^{*m}(x_n + \frac{\sqrt{k_n}}{nf(x_n)} \cdot u) du \\ & \leq 2m c_1 \frac{\sqrt{k_n}}{n} \cdot (t-s) \end{aligned}$$

where  $c_1 > 0$  is some absolute constant.

For  $\delta > 0$  and integer  $r$  to be specified later, define

$$\begin{aligned} Q_{n,r}(s) &:= \frac{n}{\sqrt{k_n}} \frac{1}{\binom{n}{m}} \sum_c \{h(x_{i_1}, x_{i_2}, \dots, x_{i_m}; x_n + \frac{\sqrt{k_n}}{nf(x_n)} \cdot s) \\ & \quad - h(x_{i_1}, x_{i_2}, \dots, x_{i_m}; x_n + \frac{\sqrt{k_n}}{nf(x_n)} \cdot r\delta)\}. \end{aligned}$$

Then

$$(4.12) \quad Q_n(s) = Q_{n,r}(s) + Q_{n,0}(r\delta).$$



First consider  $Q_{n,0}(r\delta)$ . Then

$$EQ_{n,0}^2(r\delta) = \frac{n^2}{k_n} \frac{1}{\left[\frac{n}{m}\right]^2} \sum_c \sum_c E\{h(X_{i_1}, \dots, X_{i_m}; x_n + \frac{\sqrt{k_n}}{nf(x_n)} r\delta) - h(X_{i_1}, \dots, X_{i_m}; x_n)\} \cdot \\ \{h(X_{j_1}, \dots, X_{j_m}; x_n + \frac{\sqrt{k_n}}{nf(x_n)} r\delta) - h(X_{j_1}, \dots, X_{j_m}; x_n)\}.$$

Consider all terms with  $\ell \leq m$  equal components, then by the boundedness of  $h$  and (4.11), the contribution of these terms will be smaller than

$$c_2 \frac{n^2}{k_n} \frac{1}{\left[\frac{n}{m}\right]^2} \binom{n}{2m-\ell} \frac{\sqrt{k_n}}{n} r\delta \sim c_2 n^{1-\ell} \frac{\delta r}{\sqrt{k_n}} \quad (n \rightarrow \infty).$$

If  $\ell = 0$ , obviously the expectation of the product is zero, so we may write

$$EQ_{n,0}^2(r\delta) \leq c_3 \cdot \frac{\delta r}{\sqrt{k_n}}$$

implying that

$$(4.13) \quad Q_{n,0}(r\delta) \xrightarrow{p} 0 \quad \text{as } n \rightarrow \infty.$$

We now treat  $Q_{n,r}(s)$ .

Denote

$$H_{r,n}(X_{i_1}, \dots, X_{i_m}) = I(x_n + \frac{\sqrt{k_n}}{nf(x_n)} r\delta \leq X_{i_1} + \dots + X_{i_m} \leq x_n + \frac{\sqrt{k_n}}{nf(x_n)} (r+1)\delta) \\ + F^{*m}(x_n + \frac{\sqrt{k_n}}{nf(x_n)} (r+1)\delta) - F^{*m}(x_n + \frac{\sqrt{k_n}}{nf(x_n)} r\delta).$$

then by (4.11),

$$(4.14) \quad \sup_{r\delta \leq s < (r+1)\delta} |Q_{n,r}(s)| \leq \frac{n}{\sqrt{k_n}} \frac{1}{\left[\frac{n}{m}\right]} \sum_c H_{r,n}(X_{i_1}, \dots, X_{i_m}) \\ \leq D_n + 2mc_1 \cdot \delta$$

where

$$D_n = \frac{n}{\sqrt{k_n}} \cdot \frac{1}{\left[\frac{n}{m}\right]} \sum_c \{H_{r,n}(X_{i_1}, \dots, X_{i_m}) - E H_{r,n}(X_{i_1}, \dots, X_{i_m})\}$$

In the same way as for  $Q_{n,0}(r\delta)$ , one can show that  $ED_n^2 \rightarrow 0$  so that  $D_n \xrightarrow{p} 0$ .

Now let  $C$  be any bounded set in  $\mathbb{R}$  and let  $\epsilon$  be arbitrary.

Choose  $\delta = \epsilon/8mc_1$ , then  $C \subseteq \bigcup_{r \in K} [r\delta, (r+1)\delta)$  with  $K$  a finite set of integers.

By (4.12),

$$\sup_{s \in C} Q_n(s) \leq \sup_{r \in K} \left\{ \sup_{r\delta \leq s < (r+1)\delta} Q_{n,r}(s) + Q_{n,0}(r\delta) \right\}$$

such that

$$P(\sup_{s \in C} Q_n(s) > \epsilon) \leq (\#(K)) \{ P(\sup_{r\delta \leq s < (r+1)\delta} |Q_{n,r}(s)| > \epsilon/2) + P(Q_{n,0}(r\delta) > \epsilon/2) \}.$$

Then by (4.13) and (4.14),

$$\sup_{s \in C} Q_n(s) \xrightarrow{P} 0,$$

proving the lemma. □

### 5. Some comments and concluding remarks.

1. Writing the statement in theorem 4.1 in the following way

$$\frac{n}{\sqrt{k_n}} \left( \frac{k_n}{n} U_n(X_{n-k_n:n}) - \bar{F}^{*m}(x_n) \right) \xrightarrow{d} N(0, 4m^2)$$

shows that  $\frac{k_n}{n} U_n(X_{n-k_n:n})$  is a consistent and asymptotically normal estimator for the behaviour of the tail of the  $m$ -th convolution power of  $F$ .

2. It would be highly interesting for practical purposes to know when the ratio

$$\frac{\bar{F}^{*m}(x_n)}{\bar{F}(x_n)}$$

in theorem 4.1 can be replaced by its limit  $m$ . To establish this, we

need information on the difference

$$(5.1) \quad \bar{F}^{*m}(x_n) - m\bar{F}(x_n).$$

A second order theory for subexponential d.f., providing the asymptotic behaviour of (5.1), has been established by Omey and Willekens [14],[15]. Using the results in [14], we know that for a large subclass of  $\mathcal{V}$ , the difference in (5.1) behaves as  $2\mu(\frac{m}{2})f(x_n)$  ( $n \rightarrow \infty$ ), where  $\mu = \int_0^\infty x dF(x)$ . We then have the following

Corollary 5.1

Suppose that the conditions of theorem 4.1 are satisfied and that

$$\lim_{x \rightarrow \infty} \frac{\overline{F}^{*2}(x) - 2\overline{F}(x)}{\overline{F}(x)} = 2\mu < \infty.$$

If  $k_n$  is such that  $nf(x_n) = o(\sqrt{k_n})$  ( $n \rightarrow \infty$ ), then

$$\sqrt{k_n}(U_n(X_{n-k_n:n}) - m) \rightarrow N(0, 4m^2).$$

The condition on  $k_n$  in corollary 5.1 involves the density  $f$  and shows that  $k_n$  must not grow too fast as  $n \rightarrow \infty$ .

In the special situation that  $F$  has a regularly varying tail, we have from [14].

Corollary 5.2

Let  $X_1, X_2, \dots, X_n$  be a sequence of i.i.d. random variables with d.f.  $F$  such that  $\overline{F}$  is regularly varying with index  $-\alpha$ . Suppose that (2.2) and (2.3) are satisfied.

If  $k_n$  is a sequence such that  $k_n = o(n^{\frac{1}{\alpha/2 + 1}})$  and  $\log^3 n = o(k_n)$  ( $n \rightarrow \infty$ ), then

$$(5.2) \quad \sqrt{k_n}(U_n(X_{n-k_n:n}) - m) \xrightarrow{d} N(0, 4m^2)$$

Clearly, if  $k_n = O(\log^\beta n)$  for some  $\beta > 3$ , (5.2) holds uniformly over the class of d.f. with regularly varying tails which satisfy (2.2) and (2.3).

3. It is well known that the class  $\mathcal{G}$  can be embedded in the family  $\{\mathcal{G}(\gamma), \gamma \geq 0\}$ , where d.f.  $F$  in  $\mathcal{G}(\gamma)$  satisfy

$$\lim_{x \rightarrow \infty} \frac{\overline{F}^{*2}(x)}{\overline{F}(x)} = \hat{f}(-\gamma) < \infty$$

with  $\hat{f}$  the Laplace transform of  $F$ .

A similar result as in theorem 4.1 can be established for the classes  $\mathcal{G}(\gamma)$ ,

$\gamma > 0$ , but in this case the asymptotic variance will depend on  $F$ .

### Acknowledgement

Research done while the author was visiting the University of North Carolina at Chapel Hill. He likes to thank the Department of Statistics for their kind hospitality and financial support.

The author also takes pleasure in thanking N. Veraverbeke and P. Janssen for some helpful comments in an earlier stage of this research.

### References

1. Bingham, N.H., Goldie, C.M. Teugels, J.L.: Regular Variation. (Cambridge University Press, Cambridge, 1987)
2. Chistyakov, V.P.: A theorem on sums of independent positive random variables and its applications to branching processes. Theory Prob. Appl. 9 (1964), 640-648.
3. Chover, J., Ney, P. Wainger, S.: Functions of probability measures. J. d'Anal. Math. 26(1973), 255-302.
4. Chover, J., Ney, P., Wainger, S.: Degeneracy properties of subcritical branching processes. Ann. Prob. 1(1973), 663-673.
5. Chung, K.L.: A course in probability theory (2nd ed.) (Academic Press, New York, 1974)
6. Cline, D.B.H.: Convolutions of distributions with exponential and subexponential tails. (Preprint, Texas A&M University, 1985)
7. de Haan, L.: On regular variation and its applications to the weak convergence of sample extremes (Mathematical Centre Tracts, Amsterdam, 1970).
8. Embrechts, P., Goldie, C.M.: On closure and factorization properties of subexponential and related distributions. J. Austral. Math. Soc. (series A) 29(1980) 243-256.
9. Embrechts, P., Goldie, C.M.: On convolution tails. Stoch. Proc. Appl. 13 (1982) 262-278.
10. Embrechts, P., Goldie, C.M., Veraverbeke, N.: Subexponentiality and infinite divisibility. Z. Wahrscheinlichkeitstheorie verw. Geb. 49 (1979) 335-347.
11. Goldie, C.M.: Subexponential distributions and dominated-variation tails. J. Appl. Prob. 15 (1978) 440-442.
12. Grubel, R.: Asymptotic analysis in probability theory with Banach algebra techniques. (Habilitationsschrift, Universität Essen, 1984).
14. Oney, E., Willekens, E.: Second order behaviour of distributions subordinate to a distribution with finite mean. To appear: Comm. in Statist.: Stochastic Models 3, (1987).

15. Oney, E., Willekens, E.: Second order behaviour of the tail of a subordinated probability distribution. *Stoch. Proc. Appl.* 21 (1986) 339-353.
16. Pierce, D.A.: The asymptotic effect of substituting estimators for parameters in certain types of statistics. *Ann. Statist.* 10 (1982) 475-478.
17. Pitman, E.J.G.: Subexponential distribution functions. *J. Austral. Math. Soc. (Series A)* 29 (1980) 337-347.
18. Randles, R.H.: On the asymptotic normality of statistics with estimated parameters. *Ann. Statist.* 10 (1982) 462-474.
19. Seneta, E.: Regularly varying functions (Lecture Notes in Math. 508, Springer-Verlag, Berlin, 1974)
20. Serfling, R.J.: Approximation theorems of mathematical statistics. (Wiley, New York, 1980).
21. Sukhatme, B.V.: Testing the hypothesis that two populations differ only in location. *Ann. Math. Statist.* 29 (1958) 60-78.
22. Teugels, J.L.: The class of subexponential distributions. *Ann. Prob.* 3 (1975) 1000-1011.
23. Watts, V.: The almost sure representation of intermediate order statistics. *Z. Wahrscheinlichkeitstheorie verw. Geb.* 54 (1980) 281-295.
24. Willekens, E.: Higher order theory for subexponential distributions (Ph.D. thesis, K.U. Leuven, 1986).